

Reputation Deals: A Theory of Endogenous Teams

Günther Lang*

Universidade Nova de Lisboa
Faculdade de Economia[†]

first version: February 2001
this version: September 2001.

Abstract

This paper complements the traditional theory of teams (Fama (1980), Holmstrom (1982a,b)) by introducing endogenous team formation by agents who are concerned with their reputations and are informed about the types of their potential teammates. Such a situation leads to a tradeoff between joining a high-productivity type but a low-reputation partner. It is examined whether there are gains from trade, both, for the case of non-transferable and transferable utility, and what can be learned from reputation deals. Furthermore, a signaling model of teaming is developed that captures in a consistent way the process of information acquisition by the agents' strategic opponent, the market.

KEYWORDS: endogenous teams, reputation, signaling, adverse selection.

JEL-Classification: C72, C78, L14, L23.

*I gratefully acknowledge the helpful comments by Luís Cabral, Pedro Pita Barros, as well as an anonymous referee. The usual disclaimer applies.

[†] Author's address: Günther Lang, Universidade Nova de Lisboa, Faculdade de Economia, Travessa Estêvão Pinto, Campolide, 1099-032 Lisboa, Portugal, e-mail: lang@fe.unl.pt.

Introduction

In some real-life situations, formation of a team is not imposed, rather a voluntary act. One may think, for example, of researchers who join each other in writing papers, or legal counsellors who found law firms in order to work together. There is a variety of well-known reasons why we can observe the formation of teams, although individual operation of the potential teammates is possible. These reasons can basically be resumed under the two notions of *economies of scale* and *synergy effects*: as an example for the first, sharing equipment or auxiliary staff may save on costs. As an example for the second, a traditional researcher in the area of international economics (who has ran out of ideas) might join a game theorist (who has run out of theories) in order to write a paper in which game theory is applied to international economics.¹

In the following I will argue from the perspective of *information economics* that there are additional reasons for team-formation that go beyond the conventional scale and synergy effects. Broadly speaking, teams are sometimes associated with non-observability to outsiders of the individual contributions (that even in the case of solo performance might be jammed by some stochastic variable), and these contributions may depend on individual characteristics. The following questions then arise: a) who is interested in having his characteristics concealed (revealed), b) who could be helpful in just achieving this by agreeing to team up, and c) are there gains from trade by doing so?

Whereas the answer to the first question is quite trivial, it is the second and third one that attracts the attention. If markets pay compensation according to perceived productivity, then agents are interested in achieving high reputation. Furthermore, in many situations, although reputation of an agent is common knowledge, his characteristics are private information to him — or to other agents who potentially could join him in a team. Two lawyers, for example, might know each other from law school where they had done joint course work. Hence, each one probably is well informed about the skills of the other, the market, however, after observing the initial stages of their careers, holds only imperfect information about these abilities. When then reasoning about the usefulness of forming a team with another agent, two effects need to be traded off against each other: on the one hand, having a partner with high ability increases the probability of good results — and low-ability types will be able to free-ride on this. On the other hand, choosing a partner with higher reputation than one's own will persuade a Bayesian updater to attribute responsibility for good results more to the other, while bad results are more blamed on oneself.

It is this tradeoff that motivates the analysis conducted in the present work. Formally, I will examine the question under which conditions there are gains from trade by teaming up for two agents who know each others' characteristics, both for the case of transferable and non-transferable utility. Doing so might help us to learn a lesson from the observation that a team is formed or not. Moreover, if the market itself — as strategic opponent of the two agents —

¹Or the two might take advantage of the CV synergy effect: one publication is still considered by some university departments indiscriminately as a publication for each of the co-authors.

wants to draw some inference, then one has to take into account of the potential manipulability of the signal represented by the decision of both to go together or to refrain from doing so. Introducing teaming-costs that depend on the combination of types allows us to analyze this issue in a signaling framework. Intuitively, one could imagine that those combinations that join higher productive abilities also do easier in organizing team work. Hence, we might expect that information is (partly) revealed by the teaming option.

The present work is a complement to the traditional theory of teams for the case that team-formation is an endogenous decision by the potential teammates and their types is common private information to them. It is related to the seminal research by Fama (1980) and Holmstrom (1982b) on the possibility that the market's perception of productivity may solve the moral hazard problem in a single-agent framework via reputation concerns, and to Jeon (1996), who looks at exactly the same issue, with the modification, however, that (a team of) two agents exert unobservable effort, and therefore, are able to free-ride on each other's contribution for the sake of own reputation. In fact, incentives via reputation concerns could be weakened due to the potential of the teaming-mechanism for revealing types before actual team-production takes place.

The topic discussed in the present paper is also related to the work by Farrell and Scotchmer (1988) on *partnerships*, which are endogenously formed teams with equal sharing of group revenue. In their setting, incentives to form teams arise because scale effects from the number of teammates are in place. Concerning the composition of these teams, equal revenue sharing implies that a partition of the set of all individuals takes place, where those with similar abilities are grouped together in the same partnership.

The paper is organized as follows. Section I introduces a model of individual and joint production, the latter of which in two varieties, and then discusses the Bayesian procedure used to update reputations after output is observed. Sections II and III analyze the gains from trade for the different combinations of types, given that this updating procedure is applied. In Section IV we discuss the information revealed to an outside observer by the decision of the agents to team up or to refrain from doing so. Section V looks at a simple signaling framework that fully internalizes the strategic incentives for teaming, allowing for a consistent modeling of information revelation to the market as strategic opponent of the agents.

I. The Framework

The output of agent i is

$$q^i = \mu^i + \epsilon_i, \quad (1)$$

where μ^i is the deterministic component that, at the same time, represents i 's type, and ϵ_i is a stochastic factor that influences i 's work. As a matter of analytical tractability, we assume that both components are from the set $\{0, 1\}$. The probability of experiencing a favourable result from nature's choice is $\text{prob}(\epsilon_i = 1) = r$. We assume that the stochastic components are independently

distributed over all agents. Agent i is aware of his own type, but the market is not: all potential principals only know that agents are drawn from an a-priori-distribution, with $\text{prob}(\mu^i = 1) = p$, and that these draws are independent. Furthermore, also the disturbances are not observable to the market.

After the course of time, observations are made by the market relating to the performance of agent i , whether he has worked on his own or joined a team. We denote by p_{t+1}^i the reputation of i for being the good type, stemming from Bayesian updating of p_t^i after observing outputs in t . Initially, at the beginning of their career in $t = 0$, all agents start with $p_0^i = p$, with time, however, their reputations may take different paths. There are two possible regimes that imply differences in what is observable to the principals.

Staying Alone

The market observes agent i 's output in t , realized according to (1), and, given his reputation p_t^i , calculates a Bayesian update $p_{t+1}^i := p_{t+1}^i(q^i|p_t^i)$. Given the choice of $\{0, 1\}$ as domain for μ^i and ϵ_i , it is straightforward that

$$p_{t+1}^i(q^i|p_t^i) = \begin{cases} 0 & \text{for } q^i = 0 \\ \frac{p_t^i(1-r)}{p_t^i(1-r)+(1-p_t^i)r} & \text{for } q^i = 1 \\ 1 & \text{for } q^i = 2, \end{cases} \quad (2)$$

as $q^i \in \{0, 1, 2\}$ are the only possible results of i 's activity.

Joining a Mate

We assume that two agents who are matched with each other in t may decide to form a team. Two different settings concerning the team production technology and observability of outcomes are considered.

Model A Each one of the agents operates independently of the other on exactly one project according to (1), and the market can observe two outputs. It is, however, impossible for outsiders to distinguish which one of the agents has caused which of the outputs, and projects are attributed with equal probability to the two team mates.² Updating the reputation of each agent then is accomplished by taking into account of only the two observed outputs q^a and q^b , and

²One example could be a tailor's shop, where clients leave their clothes for modifications or commission for new ones. Although they are probably attended by someone with expertise to take their measurements, they typically do not know who will actually execute the job, and typically it is just one person doing it. Another example could be diamond cutters: it is impossible to have one doing half of the work on a raw diamond and another one the other half. Every cutter has his own style, and mixing them up would give an ugly result. Hence, although we can be quite sure that one, and only one, of the cutters has created the artwork, we do not necessarily know which one — if they decide to team up in the sense used in this paper. From a more abstract point of view, there may exist several entities who are able to do a certain type of task, though only one of them can actually be executing a specific one. These entities could join each other in a team in order to execute a bunch of identical tasks and interact via an interface with outside parties, not allowing the latter to learn the internal distribution of these tasks on individual team-members.

so $p_{t+1}^i := p_{t+1}^i(q^a, q^b | p_t^i, p_t^j)$ for $i = 1, 2$. This update is given by:

$$\begin{aligned}
& 0 & (a) \\
& \frac{p_t^i(1-p_t^j)(1-r)^2}{p_t^1(1-p_t^2)(1-r)^2 + (1-p_t^1)(1-p_t^2)r(1-r) + (1-p_t^1)p_t^2(1-r)^2 + (1-p_t^1)(1-p_t^2)(1-r)r} & (b) \\
& \frac{p_t^i p_t^j(1-r)^2 + p_t^i(1-p_t^j)(1-r)r}{p_t^1 p_t^2(1-r)^2 + p_t^1(1-p_t^2)(1-r)r + (1-p_t^1)p_t^2 r(1-r) + (1-p_t^1)(1-p_t^2)r^2} & (c) \\
& \frac{p_t^i(1-p_t^j)r(1-r)}{p_t^1(1-p_t^2)r(1-r) + (1-p_t^1)p_t^2(1-r)r} & (d) \\
& \frac{p_t^i p_t^j r(1-r) + p_t^i(1-p_t^j)r^2 + p_t^i p_t^j(1-r)r}{p_t^1 p_t^2 r(1-r) + p_t^1(1-p_t^2)r^2 + p_t^1 p_t^2(1-r)r + (1-p_t^1)p_t^2 r^2} & (e) \\
& 1 & (f)
\end{aligned} \tag{3}$$

with $j \neq i$, where the different cases correspond to the observation of the following quantities:

- (a) $(q^a = q^b = 0)$
- (b) $(q^a = 1, q^b = 0)$ or $(q^a = 0, q^b = 1)$
- (c) $(q^a = 1, q^b = 1)$
- (d) $(q^a = 2, q^b = 0)$ or $(q^a = 0, q^b = 2)$
- (e) $(q^a = 2, q^b = 1)$ or $(q^a = 1, q^b = 2)$
- (f) $(q^a = q^b = 2)$.

Model B We assume that the market can only observe the aggregate outcome of the agents' activity, i.e. $q := q^a + q^b$.³ The Bayesian update $p_{t+1}^i := p_{t+1}^i(q | p_t^i, p_t^j)$, $i = 1, 2$, after observing q then is given by:

$$\begin{aligned}
& 0 & (a) \\
& \frac{p_t^i(1-p_t^j)(1-r)^2}{p_t^1(1-p_t^2)(1-r)^2 + (1-p_t^1)(1-p_t^2)r(1-r) + (1-p_t^1)p_t^2(1-r)^2 + (1-p_t^1)(1-p_t^2)(1-r)r} & (b) \\
& \frac{p_t^i p_t^j(1-r)^2 + p_t^i(1-p_t^j)(1-r)r + p_t^i(1-p_t^j)r(1-r)}{p_t^1 p_t^2(1-r)^2 + p_t^1(1-p_t^2)(1-r)r + p_t^1(1-p_t^2)r(1-r) + (1-p_t^1)p_t^2 r(1-r) + (1-p_t^1)p_t^2(1-r)r + (1-p_t^1)(1-p_t^2)r^2} & (c) \\
& \frac{p_t^i p_t^j r(1-r) + p_t^i(1-p_t^j)r^2 + p_t^i p_t^j(1-r)r}{p_t^1 p_t^2 r(1-r) + p_t^1(1-p_t^2)r^2 + p_t^1 p_t^2(1-r)r + (1-p_t^1)p_t^2 r^2} & (d) \\
& 1 & (e) \\
& & (4)
\end{aligned}$$

with $j \neq i$, where the different cases correspond to (a) $(q = 0)$, (b) $(q = 1)$, (c) $(q = 2)$, (d) $(q = 3)$, and (e) $(q = 4)$. Clearly, the setting of Model B is less informative than that of Model A.

In order to isolate the question of whether there are gains from teaming we suppose, as a first approach, that the market does not internalize any information potentially stemming from the (strategic) decision of the two agents to do

³One could think, for example, of two scientists who join each other in writing a paper. The overall quality of the paper may well react positively on each one's contribution, which, in turn, could be composed of skills and a disturbance term. Nevertheless, one could also imagine other technologies in this case, like best-shot or weakest-link, for instance.

so. We therefore assume that reputations are updated according to (2), on the one hand, and (3) and (4), respectively, on the other.⁴

Additional assumptions must be introduced that specify present and future payoffs and how they are related to the agents' reputations. We assume that in each period there is competition among the principals for the agents, and labour contracts are based on perceived productivity. This implies that an agent who stays alone is paid his expected output, which, given his reputation, equals $\hat{w}_t^i = p_t^i + r$. If two agents team up, then it is the team that receives compensation. In case that each works on a separate project (Model A), principals are willing to pay $\tilde{w}_t^j = 1/2p_t^1 + 1/2p_t^2 + r$ per project $j = 1, 2$. $\tilde{w}_t^1 + \tilde{w}_t^2$ is paid if a principal contracts for the aggregate output of the team (Model B). We have $\hat{w}_t^1 + \hat{w}_t^2 = \tilde{w}_t^1 + \tilde{w}_t^2$, and therefore there are no short-run gains from forming a team. In fact, in each period, remunerations paid by principals to one-member and two-member teams are just equal to the team members' aggregated reputations. Hence, the question arises as to whether two agents might be able to bolster each others' future reputations by forming a team, or, at least, guarantee that one's win outweighs the loss of the other.

We will assume for the remainder that after $t+1$ both individuals retire from their professional activity, and, when deciding whether to team up or not, each one looks at only his own expected reputation for the last period of activity.⁵

The decision to team up or abstain from doing so may well depend on the agents' types, and therefore, the market could learn from the mere observation of the agents' decision. Before discussing such an inference in *Section IV*, however, we will analyze in the following section in which cases there are gains from trade, given that the market updates reputations according to observed production results.

II. Model A: Gains from Trade

None of the agents knows in advance the realizations of the stochastic components when deciding whether to join a team or not. These realizations, together with the agents' types, determine their next-period reputations via (2) and (3), respectively. In what follows we derive expected reputations for all combinations of agents' types and for the two modes of operation. For the case of staying solo, we denote this value by $E_X(p_{t+1}^i | p_t^i)$, where $X \in \{G, B\}$ denotes the type of the agent (evidently, G (B) means $\mu^i = 1$ ($\mu^i = 0$)). In case of a team, we term it as $E_{XY}(p_{t+1}^i | p_t^i, p_t^j)$ for agent i , where i is of type X and his team mate, j , is of type Y , and $X, Y \in \{G, B\}$.

⁴The calculations that follow are rather bothersome and have been performed using DERIVE 5.0. The worksheet is available on request from the author.

⁵In a more dynamic setting, qualitative results may be different if two players interact more than once. Each player then would have to take into account of his influence on the reputation of the other because this has an effect on potential gains from trade in subsequent encounters.

Two bad types: $\mu^1 = \mu^2 = 0$

We have

$$\begin{aligned} E_B(p_{t+1}^i | p_t^i) &= (1-r)p_{t+1}^i(0 | p_t^i) + r p_{t+1}^i(1 | p_t^i) \\ &= r \frac{p_t^i(1-r)}{p_t^i(1-r) + (1-p_t^i)r} \end{aligned} \quad (5)$$

and

$$\begin{aligned} E_{BB}(p_{t+1}^i | p_t^i, p_t^j) &= \\ &\frac{1}{2}[(1-r)^2 p_{t+1}^i(0,0 | p_t^i, p_t^j) + r(1-r)p_{t+1}^i(1,0 | p_t^i, p_t^j) + \\ &\quad (1-r)r p_{t+1}^i(0,1 | p_t^i, p_t^j) + r^2 p_{t+1}^i(1,1 | p_t^i, p_t^j)] \cdot 2 \\ &= 2r(1-r) \frac{p_t^i(1-p_t^j)(1-r)^2}{p_t^1(1-p_t^2)(1-r)^2 + (1-p_t^1)(1-p_t^2)r(1-r) + (1-p_t^1)p_t^2(1-r)^2 + (1-p_t^1)(1-p_t^2)r(1-r)} + \\ &\quad r^2 \frac{p_t^i p_t^j(1-r)^2 + p_t^i(1-p_t^j)(1-r)r}{p_t^1 p_t^2(1-r)^2 + p_t^1(1-p_t^2)(1-r)r + (1-p_t^1)p_t^2 r(1-r) + (1-p_t^1)(1-p_t^2)r^2}. \end{aligned} \quad (6)$$

The difference in expected reputation from staying alone and joining a team is

$$\Delta^i := E_B(p_{t+1}^i | p_t^i) - E_{BB}(p_{t+1}^i | p_t^i, p_t^j) = \frac{p_t^i(p_t^j - p_t^i)r(r-1)^3}{[r(2p_t^1 - 1) - p_t^1][r(p_t^1(4p_t^2 - 3) - 3p_t^2 + 2) + p_t^1(1 - 2p_t^2) + p_t^2]}, \quad (8)$$

and we have

$$\Delta^i \begin{cases} = 0 & \text{for } p_t^i = 0 \text{ or } p_t^i = p_t^j \\ > 0 & \text{for } p_t^i \in (0, p_t^j) \\ < 0 & \text{for } p_t^i \in (p_t^j, 1). \end{cases} \quad (9)$$

Hence, given that side-payments are not feasible, for each of two bad types, there are strict incentives to join a team iff own reputation is strictly higher than that of the other. This, of course, implies that there cannot exist strict incentives at the same time for both of them. Therefore, if both look at their own expected reputation, it is only conceivable to find a bad-bad-type team if both have the same reputation or if one has reputation zero — and then both are just indifferent between staying solo and forming a team.⁶

On the other hand, if side-payments are feasible, then trade between both bad types is conceivable if the gain for one outweighs the losses of the other when coming together. Indeed, it follows from (8) that

$$\Delta^1 + \Delta^2 \begin{cases} = 0 & \text{for } p_t^1 = p_t^2 \\ < 0 & \text{for } p_t^1 \neq p_t^2, \end{cases} \quad (10)$$

(see Appendix A for the analytical expression for $\Delta^1 + \Delta^2$) which indicates that there are strict gains from making up a team if reputations are unequal. Figure 1 gives a graphical illustration.

⁶It might be worth to note why an agent with higher reputation than the other strictly prefers team work: if he stays alone and nature draws $\epsilon_i = 0$, then his bad type is revealed. If he is in a team, then his type is only revealed if $\epsilon_i = \epsilon_j = 0$, which is less likely. Moreover, good results are more "blamed" on him than on his low-reputation team mate, and the latter is held more responsible for bad results.

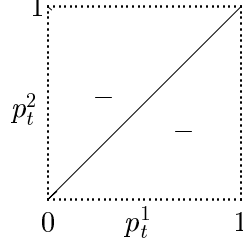


Figure 1: Two bad types. Sign of $\Delta_1 + \Delta_2$. Solid lines represent 0-loci.

Bad type, Good type: $\mu^1 = 0, \mu^2 = 1$

As before, we have

$$\begin{aligned} E_B(p_{t+1}^1 | p_t^1) &= (1-r)p_{t+1}^1(0 | p_t^1) + rp_{t+1}^1(1 | p_t^1) \\ &= r \frac{p_t^1(1-r)}{p_t^1(1-r) + (1-p_t^1)r} \end{aligned} \quad (11)$$

for the bad type. For the good type

$$\begin{aligned} E_G(p_{t+1}^2 | p_t^2) &= (1-r)p_{t+1}^2(1 | p_t^2) + rp_{t+1}^2(2 | p_t^2) \\ &= (1-r) \frac{p_t^2(1-r)}{p_t^2(1-r) + (1-p_t^2)r} + r \end{aligned} \quad (12)$$

holds.

For the bad type, joining the team yields

$$\begin{aligned} E_{BG}(p_{t+1}^1 | p_t^1, p_t^2) &= (1-r)^2 p_{t+1}^1(0, 1 | p_t^1, p_t^2) + (1-r)rp_{t+1}^1(0, 2 | p_t^1, p_t^2) + \\ &\quad r(1-r)p_{t+1}^1(1, 1 | p_t^1, p_t^2) + r^2 p_{t+1}^1(2, 1 | p_t^1, p_t^2). \end{aligned} \quad (13)$$

It follows that⁷

$$\Delta^1 := E_B(p_{t+1}^1 | p_t^1) - E_{BG}(p_{t+1}^1 | p_t^1, p_t^2) \begin{cases} = 0 & \text{for } p_t^1 = 0 \text{ or } p_t^2 = 1 \\ < 0 & \text{for } p_t^1 > 0 \text{ and } p_t^2 < 1. \end{cases} \quad (14)$$

Indeed, for $p_t^1 > 0$ and $p_t^2 < 1$, the bad type strictly prefers teaming up with a good type to staying alone: he simply knows that the good guy will guarantee higher probabilities of moderate and excellent outcomes, and he can free-ride on this — provided that he is not already known to have productivity zero ($p_t^1 = 0$) or the other is known to be of good type ($p_t^2 = 1$).

Now let's look at the perspective of the good type. His expected reputation from joining a team is

$$\begin{aligned} E_{GB}(p_{t+1}^2 | p_t^1, p_t^2) &= (1-r)^2 p_{t+1}^2(0, 1 | p_t^1, p_t^2) + (1-r)rp_{t+1}^2(0, 2 | p_t^1, p_t^2) + \\ &\quad r(1-r)p_{t+1}^2(1, 1 | p_t^1, p_t^2) + r^2 p_{t+1}^2(2, 1 | p_t^1, p_t^2) \end{aligned} \quad (15)$$

⁷Note that for $p_t^1 > 0$ and $p_t^2 < 1$ we have $E_{BG}(p_{t+1}^1 | p_t^1, p_t^2) > r(1-r)p_{t+1}^1(1, 1 | p_t^1, p_t^2) + r^2 p_{t+1}^1(2, 1 | p_t^1, p_t^2) \geq rp_{t+1}^1(1, 1 | p_t^1, p_t^2) \geq rp_{t+1}^1(1 | p_t^1) = E_B(p_{t+1}^1 | p_t^1)$.

Combining (12) and (15) we obtain⁸

$$\Delta^2 := E_G(p_{t+1}^2 | p_t^1) - E_{GB}(p_{t+1}^2 | p_t^1, p_t^2) \begin{cases} = 0 & \text{for } p_t^1 = 0 \text{ or } p_t^2 = 1 \\ > 0 & \text{for } p_t^1 > 0 \text{ and } p_t^2 < 1. \end{cases} \quad (16)$$

In fact, going together with a bad type means to share good reputation for good and intermediate results obtained from own high productivity and to share bad reputation for bad results obtained by the other's low productivity.

Now let's look at the questions whether there exist gains from trade. Figure 2 depicts graphically the sign of the generic form of $\Delta_1 + \Delta_2$, depending on p_t^1 and p_t^2 and for all ranges of r (see Appendix A for the analytical expression for $\Delta_1 + \Delta_2$).⁹

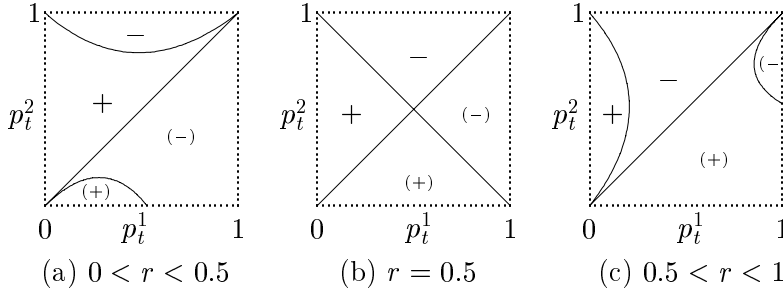


Figure 2: Bad type (1), good type (2). Sign of $\Delta_1 + \Delta_2$. Solid lines represent 0-loci. The areas indicated by (+) and (-) do only exist for $r \geq \sqrt{2} - 1$ and $r \leq 2 - \sqrt{2}$, respectively.

Two good types: $\mu^1 = \mu^2 = 1$

For each of the good types $i = 1, 2$

$$\begin{aligned} E_G(p_{t+1}^i | p_t^i) &= (1-r)p_{t+1}^2(1|p_t^i) + rp_{t+1}^i(2|p_t^i) \\ &= (1-r) \frac{p_t^i(1-r)}{p_t^i(1-r) + (1-p_t^i)r} + r, \end{aligned} \quad (17)$$

holds when staying alone.

If both join the same team, then

$$\begin{aligned} E_{GG}(p_{t+1}^i | p_t^i, p_t^j) &= (1-r)^2 p_{t+1}^i(1,1|p_t^i, p_t^j) + (1-r)rp_{t+1}^1(1,2|p_t^i, p_t^j) + \\ &\quad r(1-r)p_{t+1}^i(2,1|p_t^i, p_t^j) + r^2 p_{t+1}^i(2,2|p_t^i, p_t^j), \end{aligned} \quad (18)$$

⁸Note that for $p_t^1 > 0$ and $p_t^2 < 1$ the sum of the second and fourth term of the rhs of (15) is lower than r . At the same time, the sum of the first and third term is lower than or equal to $(1-r)p_{t+1}^2(1|p_t^2)$.

⁹Note that in Figure 2 the two areas designated by (+) and (-), respectively, do not exist over the whole range of r . The area indicated by (+) in (a) only appears for $r \geq \sqrt{2} - 1$. With r increasing, this area tends to fill the entire space below the diagonal, making the area indicated by (-) in (c) disappear for $r \geq 2 - \sqrt{2}$.

for $i = 1, 2$ and $j \neq i$.

It can be shown that

$$\Delta^i := E_G(p_{t+1}^i | p_t^i) - E_{GG}(p_{t+1}^i | p_t^i, p_t^j) \begin{cases} < 0 & \text{for } p_t^i > p_t^j \text{ and } p_t^i \neq 1 \\ = 0 & \text{for } p_t^i = p_t^j \text{ or } p_t^i = 1 \\ > 0 & \text{for } p_t^i < p_t^j \text{ and } p_t^i \neq 1, \end{cases} \quad (19)$$

which indicates that there are strict gains from joining a team iff own reputation is strictly higher than that of the other yet lower than one. This, of course, implies that there cannot be simultaneous strict gains from making up a team.

Nevertheless, the possibility of side-payments might help realize gains from trade — if they exist. One obtains, however,

$$\Delta^1 + \Delta^2 \begin{cases} = 0 & \text{for } p_t^1 = p_t^2 \\ > 0 & \text{for } p_t^1 \neq p_t^2, \end{cases} \quad (20)$$

(see Appendix A for the analytical expression for $\Delta^1 + \Delta^2$) which indicates that there are strict gains from staying alone if reputations are unequal, and if the latter are equal there is just indifference. Figure 3 gives a graphical illustration.

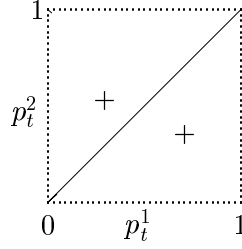


Figure 3: Two good types. Sign of $\Delta_1 + \Delta_2$. Solid lines represent 0-loci.

III. Model B: Gains from Trade

In what follows we will analyze the case where only the aggregate output can be used by the market to update reputations. Expected reputations for the different types in case of individual operation can be taken from the preceding section.

Two bad types: $\mu^1 = \mu^2 = 0$

We have

$$\begin{aligned} E_{BB}(p_{t+1}^i | p_t^i, p_t^j) &= (1-r)^2 0 + r(1-r)p_{t+1}^i(1 | p_t^i, p_t^j) \\ &\quad + (1-r)r p_{t+1}^i(1 | p_t^i, p_t^j) + r^2 p_{t+1}^i(2 | p_t^i, p_t^j) \end{aligned} \quad (21)$$

The explicit expression for the relative advantage from staying alone $\Delta^i := E_B(p_{t+1}^i | p_t^i) - E_{BB}(p_{t+1}^i | p_t^i, p_t^j)$ is given in Appendix B. Figure 4 (a) illustrates the sign of Δ^1 .

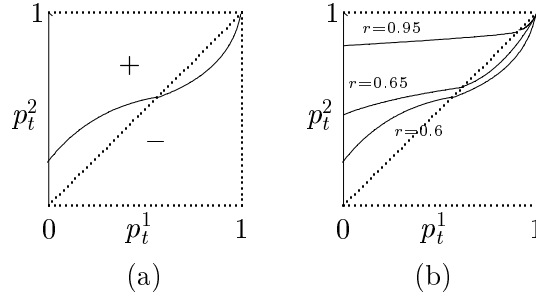


Figure 4: Two bad types. (a) Sign of Δ_1 . Solid lines represent 0-loci. (b) For $r \rightarrow 1$ the line of interior 0-loci converges towards the upper boundary of the box.

We can compare this to Model A (see (9)) where each of the two bad types has strict incentives to join the respective other iff own reputation is higher than that of the latter. For Model B, if reputations of both are not too high, then it is of advantage to join the other even if that one's reputation is slightly higher than one's own. On the other hand, if the reputations of both are sufficiently high, then staying alone is of advantage even if the other's reputation is slightly lower than one's own. The reason for this phenomenon is that the observed outcome is less informative in Model B than in Model A: if a player has relatively low reputation, then the advantage of teaming up, namely that of enjoying higher probability to maintain one's bad type unrevealed, prevails, while, if reputation is rather high, it is the disadvantage of having to share reputation for intermediate result that has more weight.

Finally, it can be shown that for $r \rightarrow 1$, being in a team becomes advantageous for a player for a set of combinations (p_t^1, p_t^2) that converges to the entire square $[0, 1] \times [0, 1]$ (see Figure 4 (b)). It appears that the benefits of having an additional stochastic variable (the other's ϵ), on which one potentially could free-ride if its realization is equal to one, outweighs the losses of having the other player free-riding on one's own. These losses become smaller with increasing r because a result of $q = 1$ is more and more attributed to luck rather than to a good type of the player.

In contrast to Model A there can be simultaneous gains from forming a team for both players, even without the possibility of side-payments. This is the case for tuples of reputations (p_t^1, p_t^2) belonging to area A of Figure 5 (a), which is obtained from Figure 4 (a) by combining the 0-loci of Δ_1 and Δ_2 . If reputations are small enough (inferior to r) and not too different, then both win from joining each other in a team.

Looking at the case of transferable utility, Figure 5 (b) depicts the sign of $\Delta_1 + \Delta_2$. Comparing this to Model A (see Figure 1), where there are gains from forming a team whenever reputations are unequal, one realizes that for high enough and not too different reputations, staying alone is of advantage.

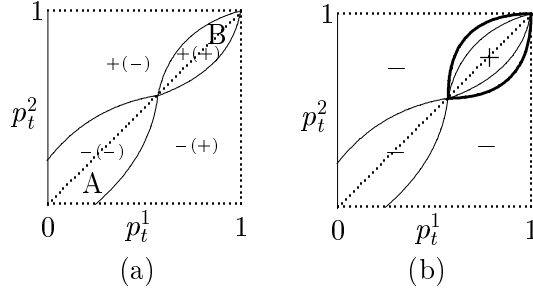


Figure 5: Two bad types. (a) Signs of Δ_1 and Δ_2 (in brackets). Solid lines represent 0-loci. (b) Sign of $\Delta_1 + \Delta_2$. Bold solid lines represent 0-loci. The other solid lines stem from (a).

Bad type, Good type: $\mu^1 = 0, \mu^2 = 1$

We have

$$E_{BG}(p_{t+1}^1 | p_t^1, p_t^2) = (1-r)^2 p_{t+1}^1(1 | p_t^1, p_t^2) + r(1-r)p_{t+1}^1(2 | p_t^1, p_t^2) + (1-r)r p_{t+1}^1(2 | p_t^1, p_t^2) + r^2 p_{t+1}^1(3 | p_t^1, p_t^2) \quad (22)$$

for the bad type, and

$$E_{GB}(p_{t+1}^2 | p_t^1, p_t^2) = (1-r)^2 p_{t+1}^2(1 | p_t^1, p_t^2) + r(1-r)p_{t+1}^2(2 | p_t^1, p_t^2) + (1-r)r p_{t+1}^2(2 | p_t^1, p_t^2) + r^2 p_{t+1}^2(3 | p_t^1, p_t^2) \quad (23)$$

for the good type.

It follows that¹⁰

$$\Delta^1 := E_B(p_{t+1}^1 | p_t^1) - E_{BG}(p_{t+1}^1 | p_t^1, p_t^2) \begin{cases} = 0 & \text{for } p_t^1 = 0 \\ < 0 & \text{for } p_t^1 > 0. \end{cases} \quad (24)$$

Analogously to Model A, for $p_t^1 > 0$ the bad type strictly prefers teaming up with a good type to staying alone, because he knows that the good guy will guarantee higher probabilities of moderate and excellent outcomes, and he can free-ride on this — provided that he is not already known to have productivity zero, $p_t^1 = 0$. In contrast to Model A, however, there are gains for the bad type from teaming even if $p_t^2 = 1$: because only aggregate output is observable, for the bad type free-riding on a favourable realization of the disturbance corresponding to the good type's contribution, ϵ_2 , is possible — which does not apply when the outcome of two separate projects is observed.

Clearly, for the good type there is nothing to win by teaming up because his high productivity is occluded by additional stochastic influences — ϵ_1 and μ_1 (the latter of which being stochastic only from the point of view of the outside

¹⁰Note that for $p_t^1 > 0$ we have $E_{BG}(p_{t+1}^1 | p_t^1, p_t^2) > r(1-r)p_{t+1}^1(2 | p_t^1, p_t^2) + r^2 p_{t+1}^1(3 | p_t^1, p_t^2) \geq r p_{t+1}^1(2 | p_t^1, p_t^2) \geq r p_{t+1}^1(1 | p_t^1) = E_B(p_{t+1}^1 | p_t^1)$.

observer) — and he also cannot benefit from the bad type’s zero productivity. Hence,

$$\Delta^2 := E_G(p_{t+1}^2 | p_t^2) - E_{GB}(p_{t+1}^2 | p_t^1, p_t^2) \begin{cases} = 0 & \text{for } p_t^2 = 1 \\ > 0 & \text{for } p_t^2 < 1. \end{cases} \quad (25)$$

This means that if no side-payments are feasible, both players won’t team up.

On the other hand, if the bad type can compensate a good-type partner, then it can be shown that $\Delta_1 + \Delta_2$ behaves as illustrated in Figure 6 (a).¹¹

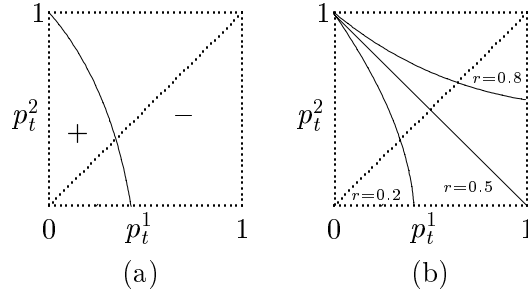


Figure 6: Bad type (1), good type (2). (a) Sign of $\Delta_1 + \Delta_2$. Solid lines represent 0-loci. (b) As $r \rightarrow 1$, the area representing gains from staying alone (+) tends monotonically to cover the entire quadrant.

Interestingly, the advantage of teaming seems to increase over some range with the good type’s reputation. This is due to the fact that his losses from joining the bad type decrease at a higher rate than the latter’s wins from joining a good-type partner. Also, as illustrated in Figure 6 (b), as r increases, the good type enjoys increasingly higher benefits from staying alone (boosting his expected reputation towards one), and this advantage cannot be compensated by the bad type’s opportunity to free-ride on ϵ_2 , that, unfortunately for him, is increasingly held responsible for contributing to relatively good results.

Two good types: $\mu^1 = \mu^2 = 1$

For both players we have

$$\begin{aligned} E_{GG}(p_{t+1}^i | p_t^i, p_t^j) &= (1-r)^2 p_{t+1}^1 (2 | p_t^i, p_t^j) + r(1-r) p_{t+1}^i (3 | p_t^i, p_t^j) \\ &\quad + (1-r)r p_{t+1}^i (3 | p_t^i, p_t^j) + r^2 p_{t+1}^i (4 | p_t^i, p_t^j), \end{aligned} \quad (26)$$

$i, j = 1, 2, j \neq i$.

In Appendix B one finds an explicit expression for Δ_i . Figure 7 (a) illustrates the sign of Δ_1 .

¹¹We omit here the analytical expression for $\Delta_1 + \Delta_2$ because of its enormous unhandiness and refer the interested reader to the DERIVE 5.0 worksheet available from the author. Note, that the curve of 0-loci intersects the diagonal at $p_t^1 = p_t^2 = r$.

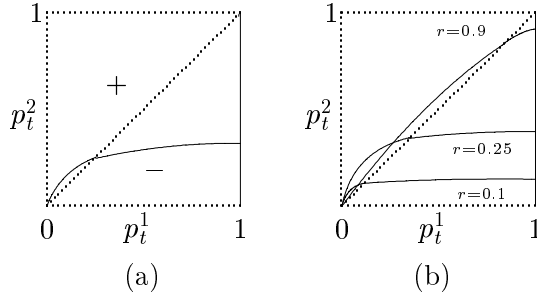


Figure 7: Two good types. (a) Sign of Δ_1 . Solid lines represent 0-loci. (b) For $r \rightarrow 1$ the line of interior 0-loci converges towards the diagonal of the box.

We can compare this to Model A (see (19)), where a good type has strict incentives to join another good type iff own reputation is strictly higher than that of the other. In contrast to this, for a relatively low reputation (lower than r) of a good-type player, he is willing to join the other good type even if that one's reputation is slightly higher (but not too high), and for relatively high reputations (higher than r), staying alone is better even if the other's reputation is slightly lower (but not too low).

In order to understand this, one must realize that Models A and B, though not equivalent, are similar, in the sense that Bayesian updating is the same if the aggregated outputs of Model A, q^a and q^b , are equal to q of Model B — except for $q^a + q^b = q = 2$. This has important implications. In particular, looking at the two good types considered here, if $\epsilon_1 = \epsilon_2 = 0$ then in Model A we observe $q^a = q^b = 1$ and in Model B $q = 2$. Clearly, $q = 2$ is less informative to outsiders than $q^a = q^b = 1$ because, unlike the latter, it cannot exclude neither $q^a = 2, q^b = 0$ nor $q^a = 0, q^b = 2$, i.e. the existence of a good-type player with good luck and a bad-type player with bad luck. This can be of advantage for either of the two good types involved, provided that the reputation of the respective other player is not too high to take all the credit. In fact, it can be shown that $E_{GG}^{(B)}(p_{t+1}^i | p_t^i, p_t^j) - E_{GG}^{(A)}(p_{t+1}^i | p_t^i, p_t^j) > (<) 0$ iff $p_t^j < (>) r$, for all $0 < p_t^i < 1$. Hence, the difference between Figure 7 (a) and (19).

We can therefore conclude that, in contrast to what has been found for Model A, there are gains from teaming for two good types even in case of non-transferable utility (see Figure 8 (a)). The gains for the case of transferable utility, $\Delta_1 + \Delta_2$, are depicted in Figure 8 (b) and are represented by a set of reputations that contains area A from (a).

IV. Informational Content of Reputation Deals

With the results of the preceding section in hands, we are in a position to analyze what can be learned by an outside observer from the fact that two agents with observable reputations join each other or abstain from doing so.

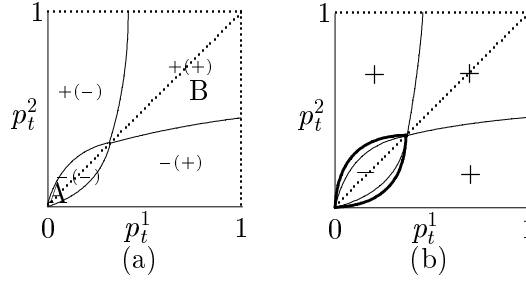


Figure 8: Two good types. (a) Signs of Δ_1 and Δ_2 (in brackets). Solid lines represent 0-loci. (b) Sign of $\Delta_1 + \Delta_2$. Bold solid lines represent 0-loci. The other solid lines stem from (a).

Model A

Non-Transferable Utility

It is clear, of course, that without the possibility of side-payments there are never strict incentives for a team to be formed. Hence, nothing can be learned concerning the types of the agents. As will become clear in the following, the same conclusion is not warranted in case of possible side-payments.

Transferable Utility

First of all, it must be observed that, if one good type and one bad type are involved, it cannot in all cases be inferred who is who. To see this, one needs to take into account that Figure 2 represents the joint gains from forming a team under the assumption that player 1 is bad type and player 2 is good type. Figure 9 below combines this case with the opposite one.

It is then clear that we cannot distinguish the two players if the aggregate relative reputation gains $\Delta_1 + \Delta_2$ come from an area where either they form a team or stay alone, independently of who is good and who is bad. This occurs in areas A,C,D and F for cases (a) and (c), and in all areas for case (b). Or, to put it the other way around: only if $r \neq 0.5$ and reputations (p_t^1, p_t^2) are from the respective areas B or E, we are able to determine who is the good guy — given that we know that one is good and the other is bad. Table 1 on p. 16 gives a resume of this result.

With this issue being clarified, we are now able to extract the information inherent in observing either a team or two individual agents doing their jobs for any given tuple of reputations (p_t^1, p_t^2) . For this purpose, we combine all the partitions stemming from Figures 1, 3 and 9. This obviously just coincides with the one of Figure 9, and so the analysis consists of looking at what can be inferred for the different areas given in this figure.

Let's look, for example, at area A for the case that $0 < r < \sqrt{2} - 1$. If we observe a team, then it is clear that only two bad types, (B,B), can have joined. Hence, both players' types are instantly revealed: $\hat{p}_t^i = 0, i = 1, 2$. On the other

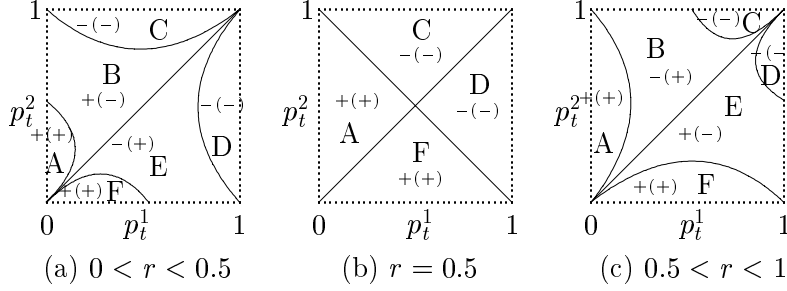


Figure 9: Bad type, good type, or vice versa. Sign of $\Delta_1 + \Delta_2$. This Figure is obtained by mirroring Figure 2 at the diagonal which starts at the origin. The signs in brackets apply for the case that player 1 is good type and player 2 bad type — the opposite to the case depicted in Figure 2. Solid lines represent 0-loci. For $r \leq \sqrt{2} - 1$ ($r \geq 2 - \sqrt{2}$) regions A and F (C and D) are not present in (a) ((c)).

hand, given that everybody stays alone, one of the other combinations of types must have occurred: (G,G), (B,G) or (G,B). Since two bad types at the same time can be excluded, both players' updated reputations, after observing their decision to stay alone, must be higher than the original ones. In fact,

$$\hat{p}_t^i = \frac{p_t^i p_t^j + p_t^i (1 - p_t^j)}{p_t^1 p_t^2 + (1 - p_t^1) p_t^2 + p_t^1 (1 - p_t^2)} > p_t^i, \quad (27)$$

$i = 1, 2, i \neq j$. Hence, if a player is member of a draw different to (B,B), then his reputation is boosted, independently of whether he is actually of good type or bad type.

Table 2 on p. 16 resumes the result of this analysis for all of the areas A-F and the whole range of r .

From this table we can gather how the different combinations of types are affected by reputation-updating from observing their decision to form a team or to stay alone. A (B,B) combination has a difficult life: In areas A and F both of them are instantly recognized as being of bad type. In areas B and E, one of them is recognized as bad, while the other stays with his original reputation. Areas C and D imply a reduction in reputation for both players, although none of them is degraded to being a bad type with certainty.

A (B,G) combination gives an advantage also to the bad type if the reputations of both are relatively low and fall in either area A or F. In areas B and E, however, either the bad type or the good type is revealed, and for areas C and D, the reputations of both will suffer.

Finally, looking at a (G,G) combination, in A and F the reputation of both players is boosted, however, not to the extent that types are revealed. However, in B or E, one of them is recognized as the good type. In C and D the good types of both are revealed.

Area	$0 < r < 0.5$		$0.5 < r < 1$	
	B	E	B	E
Team	1	2	2	1
Alone	2	1	1	2

Table 1: Who is the good guy? Given that there is one bad and one good type, information revealed in areas B and E of Figure 9 (a) and (c), respectively, from observing either a team or two agents acting solo. For all of the other areas, no conclusion can be drawn about who is who.

A	Team	(B,B)	$(-^r, -^r)$
	Alone	(G,G), (B,G), (G,B)	$(+, +)$
B	Team	(B,B), (G,B) (B,B), (B,G)	$(0, -^r)$ $(-^{\mathbf{r}}, \mathbf{0})$
	Alone	(G,G), (B,G) (G,G), (G,B)	$(0, +^r)$ $(+^{\mathbf{r}}, \mathbf{0})$
C	Team	(B,B), (B,G), (G,B)	$(-, -)$
	Alone	(G,G)	$(+^r, +^r)$
D	Team	(B,B), (B,G), (G,B)	$(-, -)$
	Alone	(G,G)	$(+^r, +^r)$
E	Team	(B,B), (B,G) (B,B), (G,B)	$(-^r, 0)$ $(\mathbf{0}, -^{\mathbf{r}})$
	Alone	(G,G), (G,B) (G,G), (B,G)	$(+^r, 0)$ $(\mathbf{0}, +^{\mathbf{r}})$
F	Team	(B,B)	$(-^r, -^r)$
	Alone	(G,G), (G,B), (B,G)	$(+, +)$

Table 2: What can be learned from either observing a team or two individualists? For all areas those combinations of types are indicated that are consistent with the observed choice. For areas B and E there is a difference between the cases $0 < r < 0.5$ and $0.5 < r < 1$ — this is indicated in **bold** for the latter one. Note that for $r = 0.5$ only areas A,C,D and F apply. The last column indicates whether the updated reputation \hat{p}_t^i is greater (+), equal (0) or lower (-) than p_t^i . The superscript r indicates when an individual's type is instantly revealed.

Model B

Non-Transferable Utility

In contrast to Model A, there are gains from teaming even in case of non-transferable utility. For relatively low reputations (lower than r) of both players, (B,B) as well as (G,G) combinations tend to team up, whereas (B,G) or (G,B) never do. We can combine Figures 4 (a) and 8 (a) in order to analyze what can be learned when observing a team. The situation in the relevant region, $[0, r] \times [0, r]$, is illustrated in Figure 10.¹²

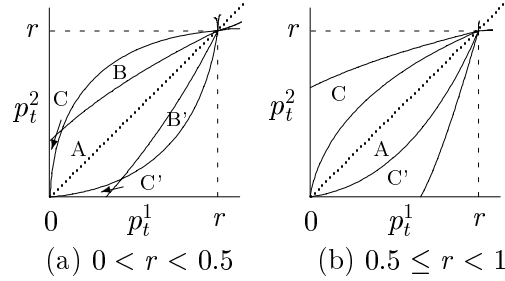


Figure 10: Case of non-transferable utility. Combinations of types who agree to team in the different areas. A: (B,B) and (G,G). B,B': (G,G). C,C': (B,B). In all other regions of the interval $[0, 1] \times [0, 1]$ at least one player prefers to stay alone.

Table 3 on p. 18 describes for all regions of $[0, 1] \times [0, 1]$ what combinations of types can be expected when observing a team or individual activity, and it compares the corresponding conditional probability of a player being of good type, \hat{p}_t^i , to the ex ante one, p_t^i .

Clearly, learning a lesson from these observations in some cases accelerates the revelation of types, in others it holds up the process. For instance, although reputations are relatively low, a (G,G) combination is instantly identified in areas B and B'. On the other hand, each of the members of a (B,B) combination receives a boost in area A, provided that $p_t^1, p_t^2 > 1/2$.

Transferable Utility

We can combine Figures 5 (b), 6 (a) and 8 (b) in order to be able to analyse what can be learned from either observing a team or individual activity. Figure 11 on p. 20 depicts the relevant partition of the unit square, and Table 4 on p. 19 presents the results.

¹²Recall that the curves in this figure are given by (42) and (44) of Appendix B. Note that for $0 < r < 0.5$ and $0.5 \leq r < 1$ two different cases emerge, with areas B and B' disappearing as r passes the barrier.

A	Team	(B,B),(G,G)	$(+, +)$ iff $p_t^1, p_t^2 > 1/2$ $(+, 0)$ iff $p_t^1 = 1/2, p_t^2 > 1/2$ $(0, +)$ iff $p_t^1 > 1/2, p_t^2 = 1/2$ $(-, 0)$ iff $p_t^1 = 1/2, p_t^2 < 1/2$ $(0, -)$ iff $p_t^1 < 1/2, p_t^2 = 1/2$ $(-, -)$ iff $p_t^1 < 1/2, p_t^2 < 1/2$
	Alone	(B,G),(G,B)	$(-, -)$ iff $p_t^1, p_t^2 > 1/2$ $(-, 0)$ iff $p_t^1 = 1/2, p_t^2 > 1/2$ $(0, -)$ iff $p_t^1 > 1/2, p_t^2 = 1/2$ $(+, 0)$ iff $p_t^1 = 1/2, p_t^2 < 1/2$ $(0, +)$ iff $p_t^1 < 1/2, p_t^2 = 1/2$ $(+, +)$ iff $p_t^1 < 1/2, p_t^2 < 1/2$
B, B'	Team	(G,G)	$(+^r, +^r)$
	Alone	(B,B),(B,G) (G,B)	$(-, -)$
C, C'	Team	(B,B)	$(-^r, -^r)$
	Alone	(B,G),(G,B) (G,G)	$(+, +)$
$[0, 1] \times [0, 1]$ less the above sets	Team	None	—
	Alone	(B,B),(B,G) (G,B),(G,G)	$(0, 0)$

Table 3: Non-transferable utility. What can be learned from either observing a team or two individualists? For all areas those combinations of types are indicated that are consistent with the observed choice. The last column indicates whether the updated reputation \hat{p}_t^i is greater (+), equal (0) or lower (-) than p_t^i . The superscript r indicates when an individual's type is instantly revealed.

A E	Team Alone	(B,B),(G,G)	$(+, +)$ iff $p_t^1, p_t^2 > 1/2$ $(+, 0)$ iff $p_t^1 = 1/2, p_t^2 > 1/2$ $(0, +)$ iff $p_t^1 > 1/2, p_t^2 = 1/2$ $(-, 0)$ iff $p_t^1 = 1/2, p_t^2 < 1/2$ $(0, -)$ iff $p_t^1 < 1/2, p_t^2 = 1/2$ $(-, -)$ iff $p_t^1 < 1/2, p_t^2 < 1/2$
	Alone Team	(B,G),(G,B)	$(-, -)$ iff $p_t^1, p_t^2 > 1/2$ $(-, 0)$ iff $p_t^1 = 1/2, p_t^2 > 1/2$ $(0, -)$ iff $p_t^1 > 1/2, p_t^2 = 1/2$ $(+, 0)$ iff $p_t^1 = 1/2, p_t^2 < 1/2$ $(0, +)$ iff $p_t^1 < 1/2, p_t^2 = 1/2$ $(+, +)$ iff $p_t^1 < 1/2, p_t^2 < 1/2$
B	Team	(B,B)	$(-^r, -^r)$
	Alone	(B,G),(G,B) (G,G)	$(+, +)$
C	Team	(B,B),(G,B)	$(0, -^r)$
	Alone	(B,G),(G,G)	$(0, +^r)$
C'	Team	(B,B),(B,G)	$(-^r, 0)$
	Alone	(G,B),(G,G)	$(+^r, 0)$
D	Team	(B,B),(B,G) (G,B)	$(-^r, -^r)$
	Alone	(G,G)	$(+^r, +^r)$

Table 4: Transferable utility. What can be learned from either observing a team or two individualists? For all areas those combinations of types are indicated that are consistent with the observed choice. The last column indicates whether the updated reputation \hat{p}_t^i is greater (+), equal (0) or lower (-) than p_t^i . The superscript r indicates when an individual's type is instantly revealed.

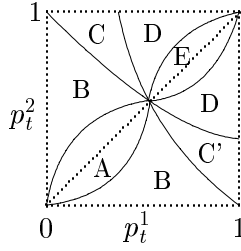


Figure 11: Transferable utility. Relevant partition of the unit square.

V. A Signaling-Theory of Team-Formation

In the setting of the previous section, the issue of learning raises the problem of strategic behaviour by the agents: if, for example, for a certain combination of reputations it is known that only good types could possibly win joining each other in a team, and if compensation were based on learning this fact, then also the other combinations of types would team up. Therefore, it is impossible to learn if agents expect us to do so — at least if we learn in this naive manner. This point, however, is based on the assumption maintained so far that teaming bears no cost. In the following we will admit costs for forming teams and develop a simple model in the tradition of signaling-theory that captures the possibility for information being (partly) revealed in a consistent way.

We assume

$$c(B, B) > c(G, B) = c(B, G) > c(G, G) \geq 0, \quad (28)$$

i.e. making up a team is most costly for the (B,B) combination and least costly for the (G,G) combination, with costs for ((G,B),(B,G)) lying in the middle. A possible interpretation could be that more capable individuals do not only exhibit higher performance concerning their productive activity, but do also easier in organizing their professional partnerships.

In what follows we do not model negotiation between the two agents, rather we assume that they team up iff there is a surplus from doing so.¹³ Clearly, it seems unreasonable to assume that the process of negotiation could be observed publicly, and that potentially lessons could already be learned by interpreting its course and result. We therefore abstract from the internal interaction between the two agents and assume the existence of two players: a) the union of the two potential teammates, on the one hand, and b) the market, on the other.

We are then in a signaling framework in the tradition of Spence (1973), and in what follows we will work out the perfect Bayesian equilibria in pure strategies (see Fudenberg and Tirole (1991a,b), Gibbons (1992)) that may exist.

¹³Note that it does not make any difference whether we consider the production technology of Model A or B, because in both cases the value of aggregated output is the same for any combination of types.

No teams at all

This cases is associated with

$$2 - c(G, G) < 0. \quad (29)$$

This means that for a (G,G) combination, teaming up never pays off, even if it is believed that only (G,G)'s join each other. Obviously, because of (28), also neither of the other combinations of types would then form a team. Therefore, only pooling equilibria in which all stay solo do exist.

Only (G,G) potentially team up

This case is associated with

$$\begin{aligned} 2 - c(G, G) &\geq 0 \\ 2 - c(G, B) &< 0. \end{aligned}$$

Clearly, neither of the combinations (B,B),(G,B),(B,G) would ever form a team. Hence, for a pooling equilibrium, beliefs out-of-equilibrium at the information set where a team is observed must attribute probability 1 to the (G,G) combination. Then an equilibrium where no combination teams up exists if

$$2 - c(G, G) \leq p_t^1 + p_t^2. \quad (30)$$

On the other hand, (G,G) separate from the other combinations by forming a team if

$$2 - c(G, G) \geq \hat{p}_t^1 + \hat{p}_t^2, \quad (31)$$

where

$$\hat{p}_t^i := \frac{p_t^i(1 - p_t^j)}{p_t^1(1 - p_t^2) + (1 - p_t^1)p_t^2 + (1 - p_t^1)(1 - p_t^2)}, i = 1, 2, j \neq i, \quad (32)$$

represents the probability of player i being a G-type, given that (B,B),(G,B) and (B,G) join each other in not teaming up. Figure 12 below illustrates the occurrence of both types of equilibria, depending on $2 - c(G, G)$.

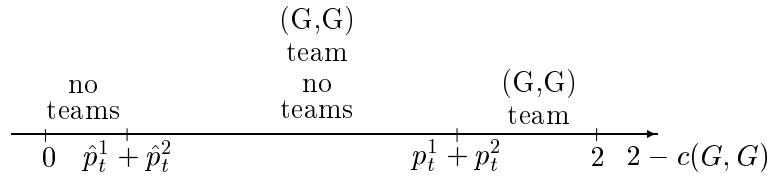


Figure 12: Occurrence of pooling and separating equilibria.

Only (G,G),((G,B) and (B,G)) potentially team up

This case is associated with

$$\begin{aligned} 2 - c(G, G) &> 0 \\ 2 - c(G, B) &\geq 0 \\ 2 - c(B, B) &< 0. \end{aligned}$$

In this case, beliefs out-of-equilibrium at the information set where a team is observed must attribute probability 0 to the (B,B) combination. In the following we consider only symmetric equilibria in the sense that the (G,B) and the (B,G) combination behave the same way.

For a pooling equilibrium, suppose that beliefs out-of-equilibrium are that the (G,G) combination occurs with probability α and ((G,B),(B,G)) with probability $1 - \alpha$. Then

$$1 + \alpha - c(G, G) \leq p_t^1 + p_t^2 \quad (33)$$

impedes (G,G),((G,B),(B,G)) to deviate. Hence, pooling with no teams observed exists if (33) holds for some $\alpha \geq 0$.

On the other hand, only (G,G) team up, if

$$2 - c(G, G) \geq \hat{p}_t^1 + \hat{p}_t^2 \quad (34)$$

$$2 - c(G, B) \leq \hat{p}_t^1 + \hat{p}_t^2, \quad (35)$$

with \hat{p}_t^i as in (32). Finally, having (G,G),((G,B),(B,G)) teamed up requires

$$\tilde{p}_t^1 - c(G, G) \geq 0 \quad (36)$$

$$\tilde{p}_t^1 - c(G, B) \geq 0, \quad (37)$$

with

$$\tilde{p}_t^i := \frac{p_t^i p_t^j + p_t^i (1 - p_t^j)}{p_t^1 p_t^2 + p_t^1 (1 - p_t^2) + (1 - p_t^1) p_t^2}, i = 1, 2, i \neq j. \quad (38)$$

Figure 13 below illustrates the occurrence of all possible types of equilibria, depending on $c(G, G)$ and $c(G, B)$.

In analogy to the previous case, equilibria of type (i) and (ii) can coexist when $c(G, G)$ and $c(G, B)$ both lie between A and D. In case (a),¹⁴ an equilibrium of type (iii) cannot coexist with type (ii) because $c(G, B) \in [0, D]$ and $c(G, B) \in [B, C]$ are incompatible. Hence, (iii) is only compatible with (i) for $c(G, G) \in [A, B]$ and $c(G, B) \in [B, C]$. In case (b), (ii) and (iii) simultaneously exist for $c(G, B) \in [B, D]$ and $c(G, G) \in [0, B]$.

It is possible to exclude some of the pooling equilibria (i) using the *intuitive criterion* (Cho and Kreps (1987)): if $2 - c(G, B) < p_t^1 + p_t^2$ then the ((G,B),(B,G)) combinations cannot possibly improve over their equilibrium payoff by deviating to "team", while, if $2 - c(G, G) > p_t^1 + p_t^2$, the (G,G) combination could if it is believed that only (G,G) deviates. Hence, for $c(G, G) < 2 - (p_t^1 + p_t^2) < c(G, B)$, those pooling equilibria with $c(G, G)$ to the right of A and to the left of $2 - (p_t^1 + p_t^2)$ (which in turn lies to the left of B, but is not given in the figure) can be excluded.

¹⁴Note that $\tilde{p}_t^1 + \tilde{p}_t^2 \geq 2 - (\hat{p}_t^1 + \hat{p}_t^2)$ iff $p_t^1 + p_t^2 \geq 1$.

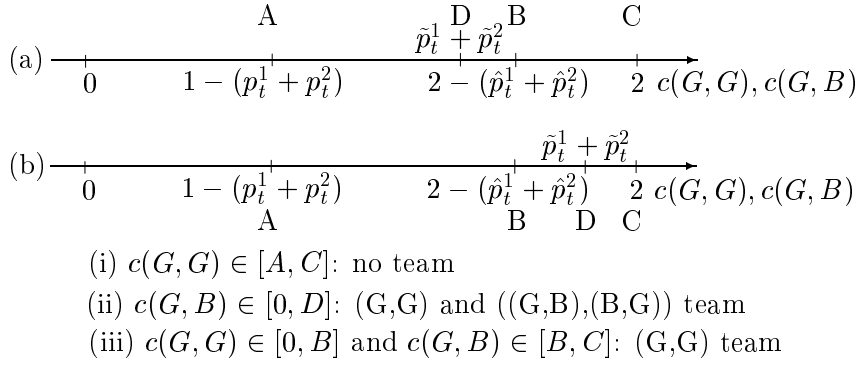


Figure 13: Occurrence of the three types of equilibria.

All potentially team up

This case is associated with

$$\begin{aligned} 2 - c(G, G) &> 0 \\ 2 - c(G, B) &> 0 \\ 2 - c(B, B) &\geq 0. \end{aligned}$$

First let's look at pooling equilibria where everybody stays solo. Suppose that beliefs out-of-equilibrium are that the (G,G) combination occurs with probability α , ((G,B),(B,G)) with probability β , and with $1 - \alpha - \beta$, (B,B) shows up. Then no combination joins a team if

$$2\alpha + \beta - c(G, G) \leq p_t^1 + p_t^2 \quad (39)$$

Hence, pooling on "solo" always exists, with α and β chosen such that $\alpha + \beta \leq 1$ and (39) satisfied.

An equilibrium with only (G,G) teaming up exists if (34) and (35), together with $2 - c(B, B) < \hat{p}_t^1 + \hat{p}_t^2$ hold (whereby the latter condition is implied by (35)).

(G,G) and ((G,B),(B,G)) team up if (36) and (37) hold, and in addition $\hat{p}_t^1 + \hat{p}_t^2 - c(B, B) \leq 0$.

A new type of equilibrium is now conceivable: all combinations make up teams. Suppose that beliefs out-of-equilibrium, i.e. after observing "solo" are that the (G,G) combination occurs with probability α , ((G,B),(B,G)) with probability β , and (B,B) with probability $1 - \alpha - \beta$. Then no combination deviates from "team" iff

$$p_t^1 + p_t^2 - c(B, B) \geq 2\alpha + \beta \quad (40)$$

Hence, $p_t^1 + p_t^2 - c(B, B) \geq 0$ is necessary and sufficient for this type of equilibrium to exist, and choosing α and β such that $\alpha + \beta \leq 1$ and such that (40) is satisfied describes the set of all of them.

In terms of Figure 13, the following types of equilibria do exist:

- (i) $c(G, G) \in [0, C]$: no team
- (ii) $c(G, B) \in [0, D]$ and $c(B, B) \in [D, C]$: (G,G) and ((G,B),(B,G)) team
- (iii) $c(G, G) \in [0, B]$ and $c(G, B) \in [B, C]$: (G,G) team
- (iv) $c(B, B) \leq p_t^1 + p_t^2$: all combinations team up.

As before, it is possible to eliminate some of the pooling equilibria (i) by applying the intuitive criterion: if $c(G, G) < 2 - (p_t^1 + p_t^2) < c(G, B)$, only (G,G) could possibly benefit from a deviation, and therefore beliefs should attribute probability $\alpha = 1$ to this combination of types. Furthermore, if $c(G, G) < c(G, B) < 2 - (p_t^1 + p_t^2) < c(B, B)$ then (B,B) should have probability zero, i.e. $\alpha + \beta = 1$. Concerning the pooling equilibria of type (iv), the intuitive criterion does not impose restrictions on out-of-equilibrium beliefs, since for none of the combinations of types the maximal possible payoff (of 2) from choosing not to team up is equilibrium-dominated.

Conclusion

This paper has complemented the traditional theory of teams by introducing endogenous team-formation by agents who are informed about the types of their potential teammates. In a framework where the market pays compensation according to perceived productivity and updates reputations after observing outputs, from an individual agent's point of view, the decision to team up is driven by the tradeoffs between free-riding on the high productivity of the potential teammate — if possible — the opportunity to conceal own type — if necessary — and the advantage of joining a low-reputation partner. Teaming has been identified as beneficial for some combinations of types with certain reputations. Most notably, high-productivity individuals may be willing to join low-productivity ones, although it means a reputation-loss and requires compensation via side-payments. Such a behaviour might be different in a context where reputation as such counts more than its equivalent in money, though. It would also be interesting to analyze reputation deals in a longer-term context where agents interact more than once and, therefore, would have to take into account of the indirect effects of the decision to team on their own future reputations.

For the purpose of modeling information-revelation in a fully rational way, a full-fledged signaling model of team formation has been presented that assumes teaming-costs that are inversely related to aggregated ability of the team. Clearly, the teaming mechanism partially reveals information under some conditions, although it cannot be completely revealing due to instrument insufficiency. Nevertheless, in case that one considers moral hazard as an additional problem, this finding emphasizes the importance of direct incentives provided by (team-)compensation that is conditioned on the actual performance of a team (Holmstrom (1982a)). It would be interesting to look at this latter issue in a combined model, probably also considering other production technologies.

Appendix A

Gains from trade for Model A for all combinations of types, depending on initial reputations, p_t^1 and p_t^2 , and r .

Two bad types

$$\Delta^1 + \Delta^2 = \frac{r^2(p_t^1 - p_t^2)^2(r-1)^3}{(p_t^1(2r-1)-r)(p_t^1(2p_t^2(2r-1)-3r+1)+p_t^2(1-3r)+2r)(p_t^2(2r-1)-r)}$$

Bad type, Good type

$$\begin{aligned} \Delta^1 + \Delta^2 = & \frac{r(1-r)(p_t^1 - p_t^2) \cdot (p_t^2)^2(2p_t^2(2r-1)^3 + p_t^2(1-2r)(11r^2 - 9r + 2))}{p_t^1(2r-1)-r)(p_t^1(2p_t^2(2r-1)-r)-p_t^2r)(p_t^1(2p_t^2(2r-1)-3r+1)+p_t^2(1-3r)+2r)(p_t^2(2r-1)-r)} \\ & + \frac{r(r^3 + 4r^2 - 4r + 1) - p_t^1r(p_t^2)^2(10r^2 - 11r + 3) - p_t^1(14r^2 - 13r + 3) + r(r^2 + 2 \cdot r - 1) - p_t^2r^2(p_t^2 - 1)(r^2 - 2r + 1))}{p_t^1(2r-1)-r)(p_t^1(2p_t^2(2r-1)-r)-p_t^2r)(p_t^1(2p_t^2(2r-1)-3r+1)+p_t^2(1-3r)+2r)(p_t^2(2r-1)-r)} \end{aligned}$$

Two good types

$$\Delta^1 + \Delta^2 = \frac{r^3(p_t^1 - p_t^2)^2(r-1)^2}{(r - p_t^2(2r-1))(p_t^1(2r-1)-r)(p_t^1(2p_t^2(2r-1)-r) - p_t^2r)}$$

Appendix B

Relative gains from staying alone for Model B for all combinations of types, depending on initial reputations, p_t^1 and p_t^2 , and r .

Two bad types

$$\begin{aligned} \Delta^i = & \{p_t^i r(1-r)(p_t^i)^2(2p_t^j)^2 r^2(2r-1) + p_t^j(2r^4 - 19r^3 + 20r^2 - 8r + 1) \\ & + r(7r^2 - 7r + 2) - p_t^i(p_t^j)^2(6r^4 - 11r^3 + 16r^2 - 8r + 1) \\ & - p_t^j r^2(7r^2 + 2r - 1) + r^2(4r^2 + r - 1)) \\ & + r(p_t^j)^2(3r^3 - 5r^2 + 6r - 2) - p_t^j r(4r^2 - r + 1) \\ & + 2r^3)\} / \{(p_t^i(2r-1)-r)(p_t^i(2p_t^j(2r-1)-3r+1) \\ & + p_t^j(1-3r)+2r)(p_t^i(p_t^j(6r^2-6r+1) \\ & - r(3r-2)) - r(p_t^j(3r-2)-r))\}, \end{aligned} \tag{41}$$

where the interior 0-loci of Δ^i are given by the following function:

$$\begin{aligned}
p_t^j(p_t^i, r) = & \{ (p_t^i)^4 (4r^6 - 68r^5 + 189r^4 - 178r^3 + 75r^2 - 14r + 1) + 2p_t^i{}^3 r (14r^5 \\
& + 15r^4 - 130r^3 + 118r^2 - 39r + 4) - p_t^i{}^2 r^2 (63r^4 - 94r^3 - 91r^2 \\
& + 86r - 18) + 2p_t^i r^3 (20r^3 - 39r^2 - 3r + 4) \\
& - r^4 (8r^2 - 16r - 1) (1 - r) - p_t^i{}^2 (2r^4 - 19r^3 + 20r^2 - 8r + 1) \\
& - r^2 (p_t^i (7r^2 + 2r - 1) - 4r^2 + r - 1) \}^{1/2} / \{ (2(2p_t^i{}^2 r^2 (2r - 1) \\
& - p_t^i (6r^4 - 11r^3 + 16r^2 - 8r + 1) + r(3r^3 - 5r^2 + 6r - 2))) \}, \tag{42}
\end{aligned}$$

$i, j = 1, 2, j \neq i$. This function intersects once with the diagonal of the space (p_t^i, p_t^j) at $p_t^i = r$. Furthermore, the corner solutions for $\Delta_i = 0$ are $p_t^i = 0$ and $p_t^i = p_t^j = 1$.

Bad type, Good type

Due to complexity not given.

Two good types

$$\begin{aligned}
\Delta^i = & r(1-r)(p_t^i - 1)(p_t^i{}^2 (2p_t^j{}^2 (r^2 - 2r + 1)(2r - 1) + p_t^j r(2r^3 + 3r^2 - 5r + 1) \\
& - r^2 (2r^2 - 1)) - p_t^i r(p_t^j{}^2 (2r - 1)(3r^2 - r - 1) - p_t^j r(r^2 - 2r + 1) \\
& - r^3) + p_t^j r^3 (p_t^j (3r - 2) - r)) / ((p_t^i (2r - 1) - r)(p_t^i (2p_t^j (2r \\
& - 1) - r) - p_t^j r)(p_t^i (p_t^j (6r^2 - 6r + 1) - r(3r - 2)) - r(p_t^j (3r - 2) - r))) \tag{43}
\end{aligned}$$

where the interior 0-loci of Δ^i are given by the following function:

$$\begin{aligned}
p_t^j(p_t^i, r) = & -r((p_t^i)^4 (4r^6 + 44r^5 - 91r^4 + 22r^3 + 55r^2 - 42r + 9) \\
& - 2p_t^i{}^3 r(22r^5 - 11r^4 - 27r^3 + 16r^2 + 5r - 3) + p_t^i{}^2 r^2 (45r^4 - 46r^3 + 6r + 1) \\
& - 2p_t^i r^4 (7r^2 - 6r + 1) + r^6) + p_t^i{}^2 (2r^3 + 3r^2 \\
& - 5r + 1) + p_t^i r(r^2 - 2r + 1) - r^3)^{1/2} / (2(2p_t^i{}^2 (2r^3 \\
& - 5r^2 + 4r - 1) - p_t^i r(6r^3 - 5r^2 - r + 1) + r^3 (3r - 2))) \tag{44}
\end{aligned}$$

$i = 1, 2, j \neq i$. This function intersects once with the diagonal of the space (p_t^i, p_t^j) at $p_t^i = r$. Furthermore, the corner solutions for $\Delta_i = 0$ are $p_t^i = 1$ and $p_t^i = p_t^j = 0$.

References

- [1] **Cho, I.-K., Kreps, D. (1987):** Signalling Games and Stable Equilibria, *Quarterly Journal of Economics*, 102, 179–222.
- [2] **Fama, E. (1980):** Agency Problems and the Theory of the Firm, *Journal of Political Economy*, 88, 288–307.
- [3] **Farrell, J. and Scotchmer, S. (1988):** Partnerships, *Quarterly Journal of Economics*, 103, 279–297.
- [4] **Fudenberg, D. and Tirole, J. (1991a):** Perfect Bayesian Equilibrium and Sequential Equilibrium, *Journal of Economic Theory*, 53, 236–260.
- [5] **Fudenberg, D. and Tirole, J. (1991b):** *Game Theory*, MIT Press, Cambridge (Mass.).
- [6] **Gibbons, R. (1992):** *A Primer in Game Theory*, Harvester Wheatsheaf, Hertfordshire.
- [7] **Holmstrom, B. (1982a):** Moral Hazard in Teams, *Bell Journal of Economics*, 10, 324–340.
- [8] **Holmstrom, B. (1982b):** Managerial Incentive Problems - a Dynamic Perspective, in: *Essays in Economics and Management in Honor of Lars Wahlbeck*, Swedish School of Economics, Helsinki.
- [9] **Jeon, S. (1996):** Moral Hazard and Reputational Concerns in Teams: Implications for Organizational Choice, *International Journal of Industrial Organization*, 14, 297–315.
- [10] **Spence, A.M. (1973):** Job Market Signalling, *Quarterly Journal of Economics*, 87, 355–374.